STUDY DESIGNS

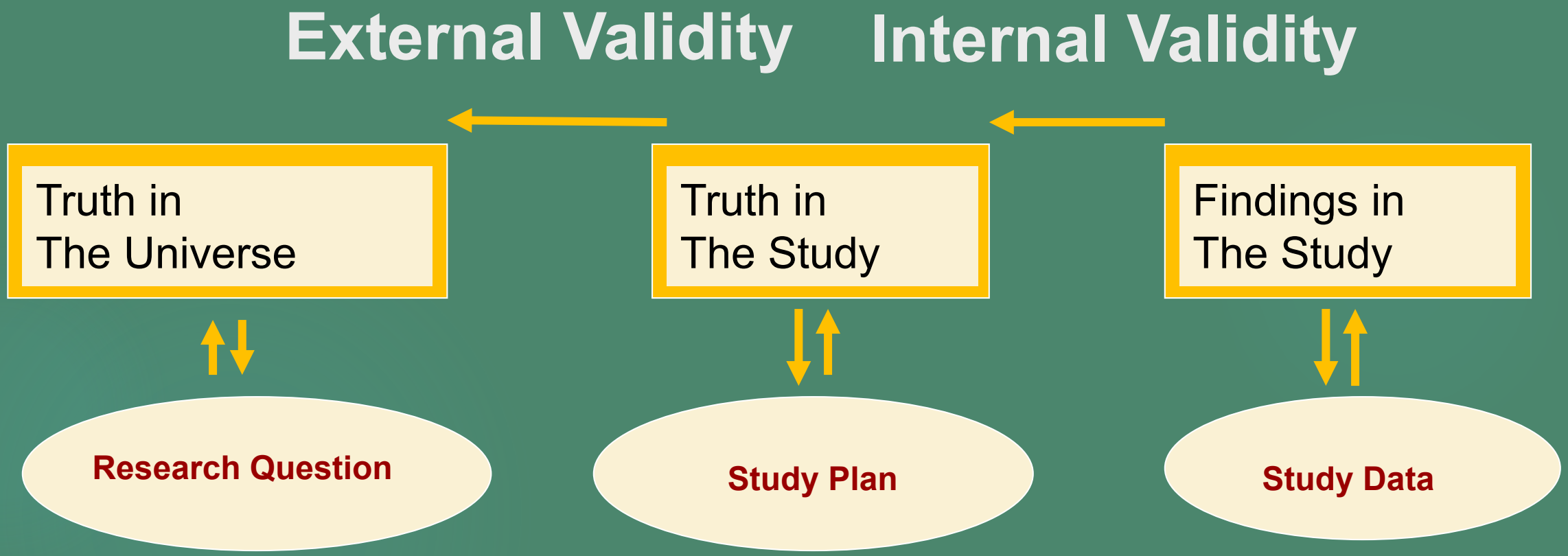
IN BIOMEDICAL RESEARCH

VALIDITY & SAMPLE SIZE

Validity is an important concept; it involves the assessment against accepted absolute standards, or in a milder form, to see if the evaluation appears to cover its intended target or targets.

INFERENCES & VALIDITIES

- Two major levels of inferences are involved in interpreting a study, a clinical trial
- The first level concerns Internal validity; the degree to which the investigator draws the correct conclusions about what actually happened in the study.
- The second level concerns External Validity (also referred to as generalizability or inference); the degree to which these conclusions could be appropriately applied to people and events outside the study.


External Validity
Internal Validity
Truth in
The Universe
Truth in
The Study
Findings in
The Study
Research Question
Study Plan
Study Data

A Simple Example:

An experiment on the effect of Vitamin C on the prevention of colds could be simply conducted as follows. A number of n children (the sample size) are randomized; half were each give a 1,000-mg tablet of Vitamin C daily during the test period and form the “experimental group”. The remaining half , who made up the “control group” received “placebo” – an identical tablet containing no Vitamin C – also on a daily basis. At the end, the “Number of colds per child” could be chosen as the outcome/response variable, and the means of the two groups are compared.

Assignment of the treatments (factor levels: Vitamin C or Placebo) to the experimental units (children) was performed using a process called "randomization". The purpose of randomization was to "balance" the characteristics of the children in each of the treatment groups, so that the difference in the response variable, the number of cold episodes per child, can be rightly attributed to the effect of the predictor – the difference between Vitamin C and Placebo.

Randomization helps to assure Internal Validity. But that is not the first step in the planning process, the “design”. In practice, we reverse the pathway; the first step is putting in efforts to assure external validity so that conclusions could be appropriately applied to people and events outside the study.The first step in the design stage is dealing with External Validity.

.

EXTERNAL VALIDITY

- Studies may be inconclusive because they were poorly planned, not enough data were collected to accomplished the goals and support the hypotheses.
- To assure external validity, we have to assure of adequate sample size (number of children in the two groups); increasing sample size will help to reduce random errors so that **conclusions could be appropriately applied to people and events outside the study**.

Of course, it is always an issue of possible trade-offs: On the one side is the issue of external validity (you need a study with large enough sample size); on the other the issue of feasibility (dictated by your ability to recruit patients). Therefore, once the study plan has been formulated, it's still a final decision: whether or not to go for it.

A TYPICAL SCENARIO

An investigator wants to randomize mice with induced tumors – say, lung tumors - into two groups; mice in one group receive placebo and the others some new agent – the effect of the new agent is to reduce the size/volume of the tumors. And he/she needs help to figure out the sample sizes.

A TYPICAL "STATISTICAL PRODUCT"

It's a statement such as:

"with 15 mice per group, we would be able to detect – with a statistical power of 80% - a reduction of 40% in tumor volume using a two-sided two-sample t-test at the 5% level"

Where do we get that 40% tumor volume specified in the Alternative Hypothesis? Whose responsibility, investigator's or statistician's?

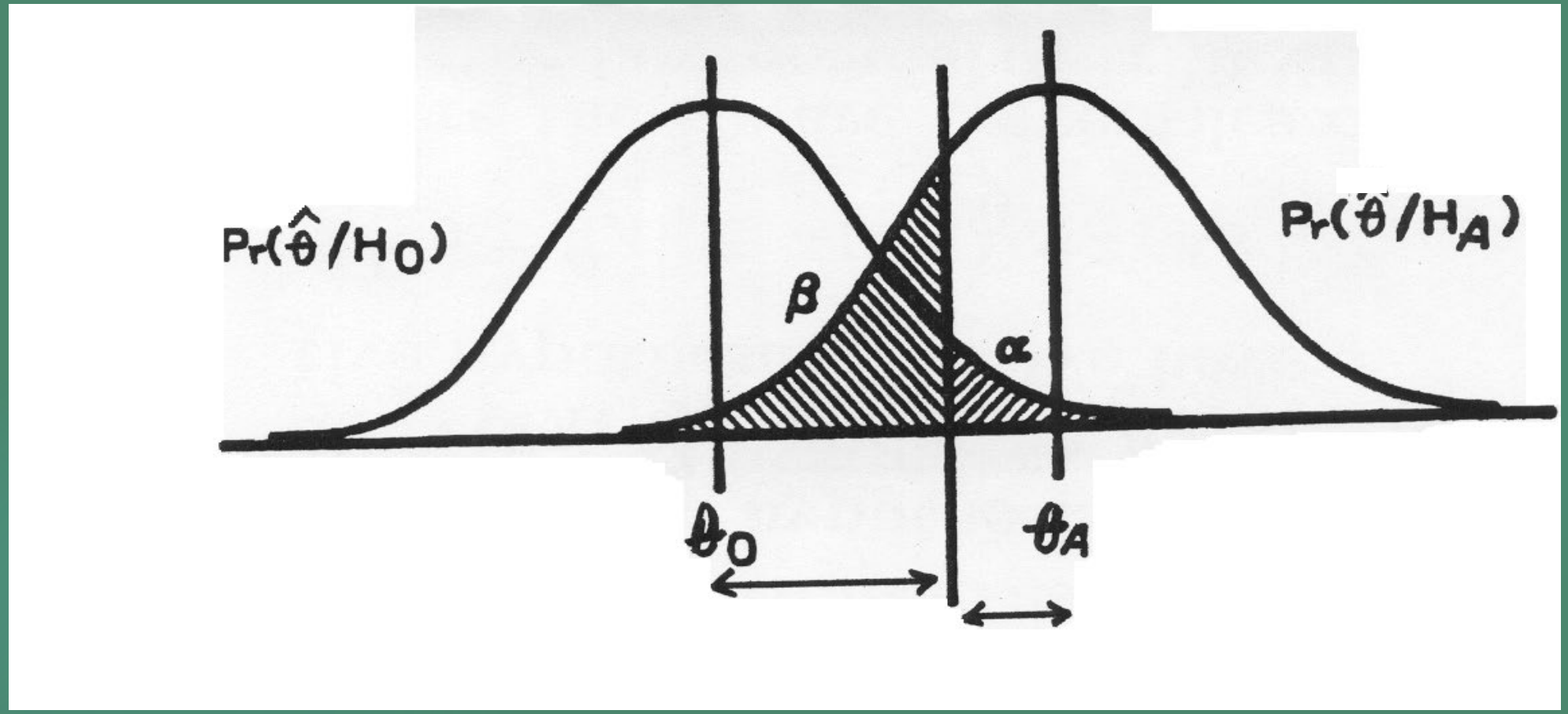
It’s a two-way need:

(1) We want to calculate a sample size for a given alternative (detectable level – maybe in the form of an “effect size”, or

(2) We need to calculate the detectable level for a given sample size

APPROACH TO SAMPLE SIZE

- The target of the investigation is a statistic θ; for example, the correlation coefficient, the odds ratio, the difference of two sample means or the difference of two sample proportions.
- Consider the statistic θ which often the MLE of some parameter (e.g. the difference of two population means), and assume that it is normally distributed as $N(\theta_0, \Sigma_0^2)$ under the null hypothesis H_0 and as $N(\theta_A, \Sigma_A^2)$ under an alternative hypothesis H_A; usually $\Sigma_0^2 = \Sigma_A^2$ or we can assume this equality for simplification.



$$| \theta_0 - \theta_A | = z_{1-\alpha}\Sigma_0 + z_{1-\beta}\Sigma_A$$

MAIN RESULT

- We have:
  $| \theta_0 - \theta_A | = z_{1-\alpha}\Sigma_0 + z_{1-\beta}\Sigma_A$
  where the z's are percentiles of N(0,1).
- Or if $\Sigma_0^2 = \Sigma_A^2 = \Sigma$, or if we assume this equality for simplification, then
  $(\theta_0 - \theta_A)^2 = (z_{1-\alpha} + z_{1-\beta})^2 \Sigma^2$
- This is the "**Basic Equation for Sample Size Determination**"; and we use $z_{1-\alpha/2}$ if the statistical test is used as two-sided.

DETECTION OF A CORRELATION

- **The Problem**: To confirm certain level of correlation between two continuously measured variables
- **The Null hypothesis** to be tested is H_0: $\rho = \rho_0$, say $\rho = 0$.
- **The Alternative hypothesis** to be tested is H_A: $\rho = \rho_A$, say $\rho = .4$.
- **The target statistic** is Pearson's "r"; indirectly through Fisher's transformation to "z".

The Coefficient of Correlation ρ between the two random variables X and Y is estimated by the (sample) Coefficient of Correlation r but the sampling distribution of r is far from being normal. Confidence intervals of is by first making the "Fisher's z transformation"; the distribution of z is normal if the sample size is not too small

$$z = \frac{1}{2} \ln\left(\frac{1+r}{1-r}\right)$$

$$z \in \text{Normal}$$

$$E(z) = \frac{1}{2} \ln\left(\frac{1+\rho}{1-\rho}\right)$$

$$\sigma^2(z) = \frac{1}{n-3}$$

RESULTS FOR CORRELATION

- The null hypothesis to be tested is $H_0: \rho = 0$
- The target statistic is Fisher's z
- Basic parameters are:

$$(\theta_0 - \theta_A)^2 = (z_{1-\alpha} + z_{1-\beta})^2 \Sigma^2$$

$$\theta_0 = \frac{1}{2}\ln\frac{1+0}{1-0} = 0;\ \theta_A = \frac{1}{2}\ln\frac{1+\rho_A}{1-\rho_A};\ \text{and}\ \Sigma^2 = \frac{1}{n-3}$$

- Result: Total required sample size:

$$n = 3 + \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{\theta_A^2}$$

- Example: If $\rho_A = .4 \rightarrow \theta_A = .424$; for 5% 2-sided & 80% power:

$$n = 3 + \frac{(1.96 + .84)^2}{.424^2} \geq 47$$

COMPARISON OF TWO MEANS

- **The Problem**: The endpoint is on a continuous scale; for example, a researcher is studying a drug which is to be used to reduce the cholesterol level in adult males aged 30 and over. Subjects are to be randomized into two groups, one receiving the new drug (group 1), and one a look-alike placebo (group 2). The response variable considered is the change in cholesterol level before and after the intervention.
- **The null hypothesis** to be tested is H_0: $\mu_2 - \mu_1 = 0$
- **The target statistic** is $\theta = \bar{x}_2 - \bar{x}_1$

DIFFERENCE OF TWO MEANS

- The null hypothesis to be tested is H_0: $\mu_1 = \mu_2$
- The target statistic is $\theta = \bar{x}_2 - \bar{x}_1$
- Basic parameters are: $\theta_0 = 0$, $\theta_A = d$, and

$$(\theta_0 - \theta_A)^2 = (z_{1-\alpha} + z_{1-\beta})^2 \Sigma^2$$

Where:

$$d^2 = (z_{1-\alpha} + z_{1-\beta})^2 \Sigma^2$$

$$\Sigma^2 = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right) = \sigma^2 \frac{4}{N}$$

RESULTS FOR TWO MEANS

- The null hypothesis to be tested is H_0: $\mu_1 = \mu_2$
- The target statistic is $\theta = \bar{x}_2 - \bar{x}_1$
- Basic parameters are: $\theta_0 = 0$, $\theta_A = d$, and

Or: $d^2 = (z_{1-\alpha} + z_{1-\beta})^2 \Sigma^2$

$$\Sigma^2 = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right) = \sigma^2 \frac{4}{N}$$

$$N = 4(z_{1-\alpha} + z_{1-\beta})^2 \frac{\sigma^2}{d^2}$$

If the two groups are planned to have different sizes, with $n_1 = pN$ and $n_2 = (1-p)N$, $(0<p<1)$; then the total sample size is:

$$N = (z_{1-\alpha} + z_{1-\beta})^2 \frac{\sigma^2}{p(1-p)d^2}$$

NEEDED COMPONENTS

- This required total sample size is affected by four factors:

(1) The size α of the test; conventionally, α =.05 is used.

(2) The desired power (1-β). This value is selected by the investigator; a power of 80% or 90% is often used.

$$N = 4(z_{1-\alpha} + z_{1-\beta})^2 \frac{\sigma^2}{d^2}$$

NEEDED COMPONENTS

(3) The quantity d, called the "minimum clinical significant difference", $d = |\mu_2 - \mu_1|$, (its determination is a clinical decision, not a statistical decision).

(4) The variance of the population. This variance σ^2 is the only quantity which is difficult to determine. The exact value is unknown; we may use information from similar studies or past studies or use some "upper bound".

$$N = 4(z_{1-\alpha} + z_{1-\beta})^2 \frac{\sigma^2}{d^2}$$

EXAMPLE

- **Specifications: Suppose a researcher is studying a drug which is used to reduce the cholesterol level in adult males aged 30 or over, and wants to test it against a placebo in a balanced randomized study. Suppose also that it is important that a reduction difference of 5 be detected (d=5). We decide to preset α =.05 and want to design a study such that its power to detect a difference between means of 5 is 95% (or β =.05). Also, the variance of cholesterol reduction (with placebo) is known to be about σ^2 = 36.**
- **Result:**

$$N = 4(1.96 + 1.65)^2 \frac{36}{5^2} = 76; \text{ or } 38 \text{ subjects in each group}$$

CROSS-OVER DESIGNS

Design:

Group 1: Period #1 (PEITC; A1) – washout – Period #2 (Placebo; B2)

Group 2: Period #1 (Placebo; B1) – washout – Period #2 (PEITC; A2)

Outcome variables:

X1 = A1 - B2; and X2 = A2 - B1

From the model:

X1 is normally distributed as $N(\alpha+\beta,\sigma^2)$

X2 is normally distributed as $N(\alpha-\beta,\sigma^2)$

Let n be the desired group size (total sample size is 2n)

TESTING FOR TREATMENT EFFECT

In testing the Null hypothesis of "no treatment effects" H0: $\alpha = 0$, one can frame it as a two-sample t-test comparing the mean of X1 versus the mean of (-X2) as seen from

X1 is normally distributed as $N(\alpha+\beta,\sigma^2)$

X2 is normally distributed as $N(\alpha-\beta,\sigma^2)$

Since:

X1 = A1 - B2; and X2 = A2 - B1

In addition to the variances of A1, A2, B1, B2 – which are usually the same – we need the covariance, say between A1 and B2. Without prior knowledge, a common practice is assuming a moderate correlation level, say 0.5

MULTIPLE-GROUP TRIALS

Sometimes we want to design a multi-group clinical trials involving a (placebo) control and k treatments, say k=2 or k=3.
After the trial is ended, the first test we would do is the One-way ANOVA F-test; and there are method for sample size estimation meeting a pre-set power for this test. However, this F-test is not the primary reason for the trial; main interest is always “pairwise comparisons”

A simple strategy would be:

(1)T o identify the number of “primary” comparisons of interest (instead of all possible pairwise comparisons) – say the comparisons of each of k treatments versus the placebo control;

(2) To follow method for two-group trial allowing for multiple comparisons; for example, dividing the size of the trial (the pre-set probability of type I errors) by the number of comparisons selected in step (1).

COMPARISON OF 2 PROPORTIONS

- The Problem: The endpoint may be on a binary scale. For example, a new vaccine will be tested in which subjects are to be randomized into two groups of equal size: a control (not immunized) group (group 1), and an experimental (immunized) group (group 2). Subjects, in both control and experimental groups, will be challenged by a certain type of bacteria and we wish to compare the infection rates.
- The null hypothesis to be tested is H_0: $\pi_2 - \pi_1 = 0$
- The target statistic is $\theta = p_2 - p_1$

DIFFERENCE OF 2 PROPORTIONS

- The null hypothesis to be tested is $H_0: \pi_1 = \pi_2$
- The target statistic is $\theta = p_2 - p_1$
- Basic parameters are: $\theta_0 = 0$, $\theta_A = d$, and approximately

$$(\theta_0 - \theta_A)^2 = (z_{1-\alpha} + z_{1-\beta})^2 \Sigma^2$$

where

$$d^2 = (z_{1-\alpha} + z_{1-\beta})^2 \Sigma^2$$

$$\Sigma^2 = \bar{\pi}(1-\bar{\pi})\left(\frac{1}{n_1} + \frac{1}{n_2}\right) = \bar{\pi}(1-\bar{\pi})\frac{4}{N}$$

RESULTS FOR 2 PROPORTIONS

- The null hypothesis to be tested is $H_0: \pi_1 = \pi_2$
- The target statistic is $\theta = p_2 - p_1$
- Basic parameters are: $\theta_0 = 0$, $\theta_A = d$, and approximately

$$d^2 = (z_{1-\alpha} + z_{1-\beta})^2 \Sigma^2$$

$$\Sigma^2 = \bar{\pi}(1-\bar{\pi})\left(\frac{1}{n_1} + \frac{1}{n_2}\right) = \bar{\pi}(1-\bar{\pi})\frac{4}{N}$$

$$N = 4(z_{1-\alpha} + z_{1-\beta})^2 \frac{\bar{\pi}(1-\bar{\pi})}{d^2}$$

NEEDED COMPONENTS

- This required total sample size is affected by four factors:
- (1) The size α of the test; conventionally, α =.05 is used.
- (2) The desired power (1-β). This value is selected by the investigator; a power of 80% or 90% is often used.

$$N = 4(z_{1-\alpha} + z_{1-\beta})^2 \frac{\bar{\pi}(1-\bar{\pi})}{d^2}$$

NEEDED COMPONENTS

(3) The quantity d, also called the "minimum clinical significant difference", $d = |\pi_2 - \pi_1|$ (its determination is a clinical decision, not a statistical decision).

(4) π is the average proportion $\bar{\pi} = (\pi_2 + \pi_1)/2$; It is obvious that the planning sample size is more difficult and a good solution requires knowledge of the scientific problem, some good idea of the magnitude of the proportions themselves.

$$N = 4(z_{1-\alpha} + z_{1-\beta})^2 \frac{\bar{\pi}(1-\bar{\pi})}{d^2}$$

EXAMPLE

- **Specifications**: Suppose we wish to conduct a clinical trial of a new therapy where the rate of successes in the control group was known to be about 5%. Further, we consider the new therapy to be superior- cost, risks, and other factors considered- if its rate of successes is about 15%. In addition, We decide to preset α =.05 and want to design a study such that its power to detect the desired difference of 15% vs. 5% is 90% (or β =.10).
- **Result**:

$$N = 4(1.96 + 1.28)^2 \frac{(.10)(.90)}{(.15 - .05)^2} = 378; \text{ or } 189 \text{ per group}$$

DESIGNING CASE-CONTROL STUDIES

- Both cohort and case-control- are comparative; the validity of the conclusions is based on a comparison.
- In a cohort study, say a clinical trial, we compare the results from the "treatment group" versus the results from the "placebo group".
- In a case-control study, we compare the "cases" versus the "controls" with respect to an exposure under investigation ("exposure" could be binary or continuous).

DIFFERENT FORMULATION

- In a cohort study, for example a two-arm clinical trial, the decision at the end is based on a "difference"; difference of two means or of two proportions. The "size" of the difference is the major criterion for sample size determination.
- In a case-control study, we compare the exposure histories of the two groups. At the end, we do not search for a difference; instead, the alternative hypothesis of a case-control study is postulated in the form of a relative risk. But the two are related.

A simple strategy is to turn the Relative Risk or Odds Ratio into the size of the “difference”, then apply the same method we use with clinical trials. Let start with the simple case of a binary risk factor.

CASE-CONTROL DESIGN FOR A BINARY RISK FACTOR

- The data analysis maybe similar to that of a Clinical Trial where we want to compare two proportions.
- However in the design stage, the alternative hypothesis is formulated in the form of a relative risk ρ. Since we cannot estimate or investigate "relative risk" using a case-control design, we would treat the given number ρ as an "odds ratio", the ratio of the odds of being exposed by a case divided by the odds of being exposed by a control.

$$\rho = \frac{\frac{\pi_1}{1-\pi_1}}{\frac{\pi_0}{1-\pi_0}}$$

CLINICAL SIGNIFICANT DIFFERENCE

- From:

$$\rho = \frac{\frac{\pi_1}{1-\pi_1}}{\frac{\pi_0}{1-\pi_0}}$$

- **We solve for the proportion for the cases, and use the previous formula for sample size applies with d = π_1 - π_0:**

$$\pi_1 = \frac{\rho\pi_0}{1+(\rho-1)\pi_0}$$

CASE-CONTROL DESIGN FOR A CONTINUOUS RISK FACTOR

- Data are analyzed using Logistic Regression
- The Model is:

$$p_x = \Pr(Y = 1 \mid X = x) = \frac{1}{1 + \exp[-(\beta_0 + \beta_1 x)]}$$

$$\text{Logit} = \ln \frac{p_x}{1 - p_x} = \beta_0 + \beta_1 x$$

Key Parameter: β_1 is the log of the Odds Ratio due to one unit increase in the value of X

BAYES' THEOREM

Recall:

$$\Pr(A \mid B) = \frac{\Pr(A \text{ and } B)}{\Pr(B)} = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B \mid A)\Pr(A) + \mathrm{PR}(B \mid \text{not } A)\Pr(\text{not } A)}$$

For example,

$$\Pr(Y = 1 \mid X = x) = \frac{\Pr(X = x \mid Y = 1)\Pr(Y = 1)}{\Pr(X = x \mid Y = 1)\Pr(Y = 1) + \Pr(X = x \mid Y = 0)\Pr(Y = 0)}$$

$$\Pr(Y = 0 \mid X = x) = \frac{\Pr(X = x \mid Y = 0)\Pr(Y = 0)}{\Pr(X = x \mid Y = 0)\Pr(Y = 0) + \Pr(X = x \mid Y = 1)\Pr(Y = 1)}$$

Take the ratio, denominators are cancelled

APPLICATION TO LOGISTIC MODEL

We use the Bayes' Rule to express the ratio of posterior probabilities as the ratio of prior probabilities times the likelihood ratio:

$$\frac{\Pr(Y=1 \mid X=x)}{\Pr(Y=0 \mid X=x)} = \frac{\Pr(X=x \mid Y=1)\Pr(Y=1)}{\Pr(X=x \mid Y=0)\Pr(Y=0)}$$

$$\frac{\Pr(Y=1 \mid X=x)}{\Pr(Y=0 \mid X=x)} = \left\{\frac{\Pr(Y=1)}{\Pr(Y=0)}\right\} \left\{\frac{\Pr(X=x \mid Y=1)}{\Pr(X=x \mid Y=0)}\right\}$$

THE LOGISTIC MODEL

$$\left\{\frac{\Pr(Y=1|X=x)}{\Pr(Y=0|X=x)}\right\}=\left\{\frac{\Pr(Y=1)}{\Pr(Y=0)}\right\}\left\{\frac{\Pr(X=x|Y=1)}{\Pr(X=x|Y=0)}\right\}$$

Taking the log of the left-hand side, we obtain the Logistic Regression Model; On the right-hand side:

the ratio of prior probabilities is a constant (with respect to x) and the likelihood ratio is the ratio of two pdf's or two densities.

NORMAL COVARIATE

- **Assume that covariate X is normally distributed**

$$\text{Logit} = \text{Constant} + \ln(\text{ratio of densities})$$

$$\text{Logit} = \text{Constant} + \left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_0}{\sigma_0^2}\right)x + \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)x^2$$

$$\text{Logit} = \text{Constant} + \left(\frac{\mu_1 - \mu_0}{\sigma^2}\right)x \quad \text{if} \quad \sigma_1^2 = \sigma_0^2 = \sigma^2$$

- **The log of the Odds Ratio associated with "one standard deviation increase in value of X" is:**

$$\ln \rho = \frac{\mu_1 - \mu_0}{\sigma}; \text{ so that } d = (\ln \rho)\sigma$$

RESULT

$$N = 4(z_{1-\alpha} + z_{1-\beta})^2 \frac{\sigma^2}{d^2}$$

$$= 4(z_{1-\alpha} + z_{1-\beta})^2 \frac{\sigma^2}{(\log \rho)^2 \sigma^2}$$

$$= \frac{4(z_{1-\alpha} + z_{1-\beta})^2}{(\log \rho)^2}$$

If number of cases and number of controls are equal

RESULT

$$N = (z_{1-\alpha} + z_{1-\beta})^2 \frac{\sigma^2}{p(1-p)d^2}$$

$$= (z_{1-\alpha} + z_{1-\beta})^2 \frac{\sigma^2}{p(1-p)(\log\rho)^2\sigma^2}$$

$$= \frac{(z_{1-\alpha} + z_{1-\beta})^2}{p(1-p)(\log\rho)^2}$$

Where p is the percent of subjects with events (cases, Y=1); 0<p<1

EXAMPLE

Suppose that an investigator is considering to design a case-control study; its aim is to investigate a potential association between coronary heart disease and serum cholesterol level. Suppose further that it is desirable to detect an odds ratio $\rho = 2.0$ for a person with cholesterol level 1 standard deviation above for the mean for his or her age group using a two-sided test with a significance level of 5% and a power of 90%. Also assuming that we plan to have the same numbers of cases and controls.

$$\alpha = .05 \rightarrow z_{1-\alpha} = 1.96$$

$$\beta = .10 \rightarrow z_{1-\beta} = 1.28$$

$$N = \frac{4}{(\rho)^2}(z_{1-\alpha} + z_{1-\beta})^2$$

$$= \frac{4}{(\ln 2)^2}(1.96 + 1.28)^2$$

$$\cong 62 \text{ subjects; } 31/\text{group}$$

Suggested Exercises:

#1 Suppose we want to compare the use of medical care by black and white teenagers. The aim is to compare the proportions of kids without physical check-ups within the last two years. Some recent survey shows that these rates for blacks and whites are 17% and 7% respectively. How large should a total sample be so that it would be able to detect such a 10% difference with a power of 90% using a statistical test at the two-sided level of significance of .01?

#2 When a patient is diagnosed as having cancer of the prostate, an important question in deciding on treatment strategy for the patient is whether or not the cancer has spread to the neighboring lymp nodes. The question is so critical in prognosis and treatment that it is customary to operate on the patient (i.e., perform a laparotomy) for the sole purpose of examining the nodes and removing tissue samples to examine under the microscope for evidence of cancer. However, certain variables that can be measured without surgery may be predictive of the nodal involvement; one of which is level of serum acid phosphatase. Suppose an investigator considers to conduct a case-control study to evaluate this possible relationship between nodal involvement (cases) and level of serum acid phosphatase. Suppose further that it is desirable to detect an odd ratio of $\theta = 1.5$ for an individual with a serum acid phosphatase level of one standard deviation above the mean for his/her age group using a two-sided test with a significance level of 5 percent and a power of 80 percent. Find the total sample size needed for using a two-sided test at the .05 level of significance.